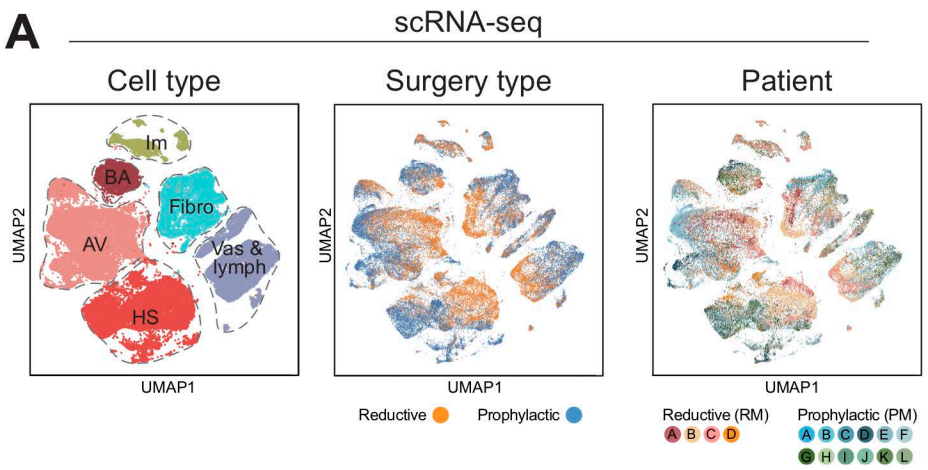
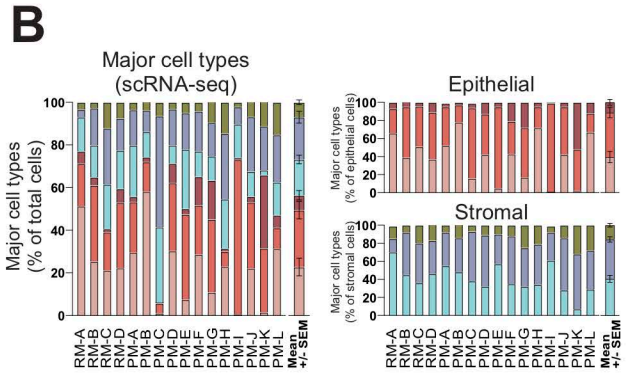
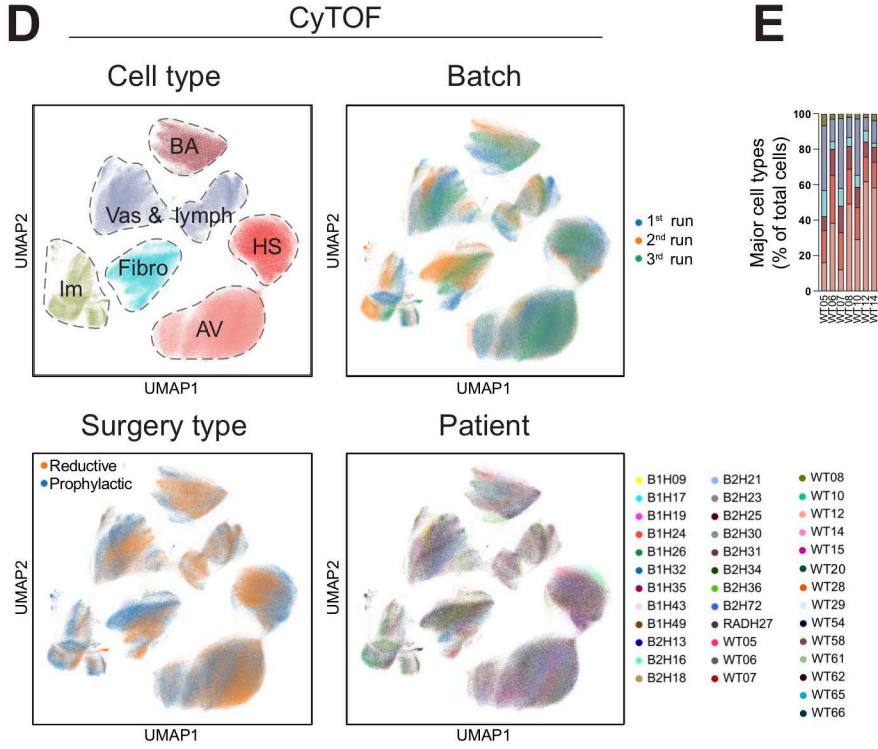
Supplemental Figure S1

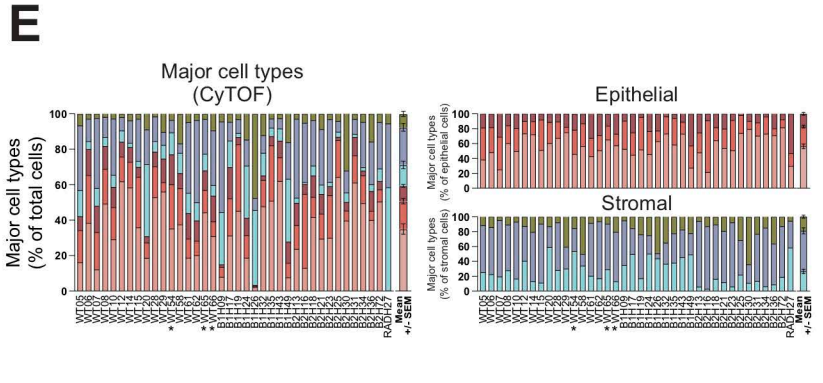




C

		Lineage markers				
		Luminal		Basal		
Protein	Gene	AV	HS	BA	Other	CyTOF runs
ANPEP	*ANPEP*					1,2,3
ANXA8	*ANXA8*					1,2,3
CD133	*PROM1*					1,2,3
CD73	*NT5E*					1,2
RANK	*TNFRSF11A*					1,2
AR	*AR*					2,3
ERα	*ESR1*					1,2,3
PRβ	*PGR*					1,2,3
GATA3	*GATA3*					1,2,3
MUC1	*MUC1*					1,2,3
HSP27	*HSPB1*					1,2,3
CD10	*MME*					1,2,3
CD90	*THY1*					1,2,3
K14	*KRT14*					1,2,3
K17	*KRT17*					1,2,3
LAM5	*LAMB3*					1,2,3
SMA	*ACTA2*					1,2,3
CD24	*CD24*					1,2,3
EPCAM	*EPCAM*					1,2,3
HER2	*ERBB2*					1,2,3
K8K18	*KRT8/18*					1,2,3
CD49F	*ITGA6*					1,2,3
EGFR	*EGFR*					1,2,3
KI67	*MKI67*				Proliferation	1,2,3
P53	*TP53*				DNA damage/stress	1,2,3
BRCA1	*BRCA1*				DNA damage/stress	1,2,3
HLA-ABC	*HLA-ABC*				MHC class I	1,2,3
H3K27me3	*N/A*				Histone modification	3
CD47	*CD47*				Other cell function	1,2,3
CD54	*ICAM1*				Other cell function	1,2,3
CD95	*FAS*				Other cell function	1,2,3
GAL1	*LGALS1*				Other cell function	1,2,3
GR	*NR3C1*				Other cell function	1,2,3
CD45	*PTPRC*				Immune cell exclusion	1,2,3
EPCR	*PROCR*				Stromal cell exclusion	1,2,3
CD140B	*PDGFRB*				Stromal cell exclusion	1,2,3
VIM	*VIM*				Stromal cell exclusion	1,2,3
CD31	*PECAM1*				Stromal cell exclusion	1,2,3
CD44	*CD44*				Stromal cell exclusion	1,2,3
C-PARP	*N/A*				Apoptotic cell exclusion	1,2,3





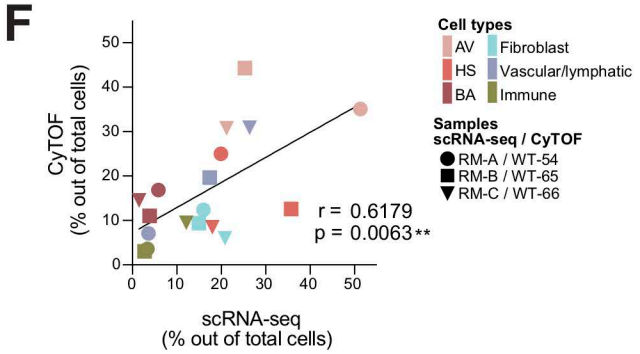
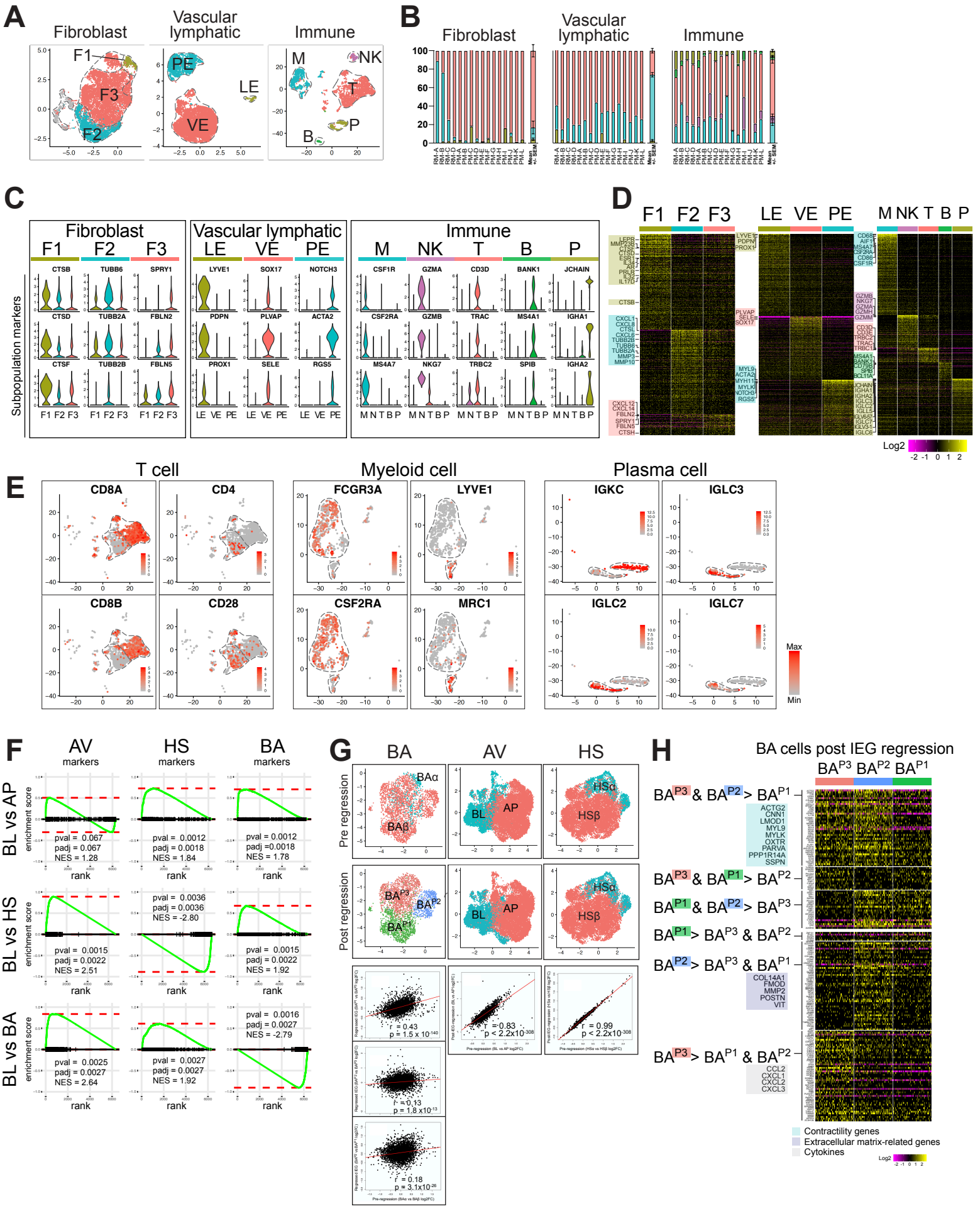
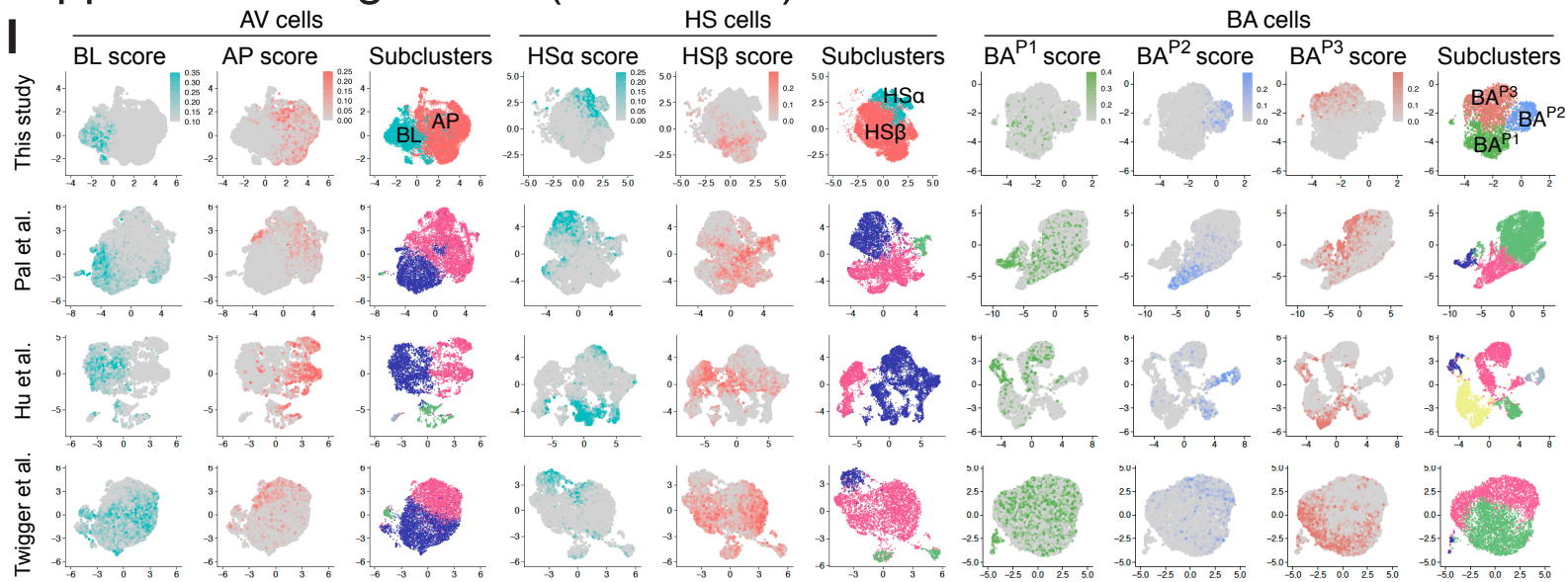


Figure S1. Major mammary epithelial and stromal cell types identified by both scRNA-seq and CyTOF
(related to Figure 1)
(A) UMAP visualization of scRNA-seq data by surgery type or donor. (See Table S1.)
(B) Proportion of the six major cell populations across 16 scRNA-seq samples. Color code is identical to (A). Reductive mammoplasty (RM) samples are noncarriers (RM-A,B,C) or *RAD51C* mutation carrier (RM-D), whereas prophylactic mastectomy (PM) samples carry mutations in *BRCA1* (PM-A through F) or *BRCA2* (PM-G through L). (See Table S1 for donor information.)
(C) CyTOF antibody panel.
(D) UMAPs of CyTOF data colored by analysis batch, surgery type, or donor. (See Table S1.)
(E) Proportion of the six major cell populations across 38 CyTOF donors annotated as noncarriers or carriers of mutations in *BRCA1* (B1), *BRCA2* (B2), or *RAD51C* (RAD). (See Table S1 for donor information.) Color code is identical to (D). CyTOF samples WT54, 65, and 66 were also analyzed by scRNA-seq as RM-A, B, and C, respectively.
(F) Correlation between scRNA-seq and CyTOF analyses of cell type proportions in three samples shared by both datasets. Pearson correlation analysis.

Supplemental Figure S2



Supplemental Figure S2 (continued)



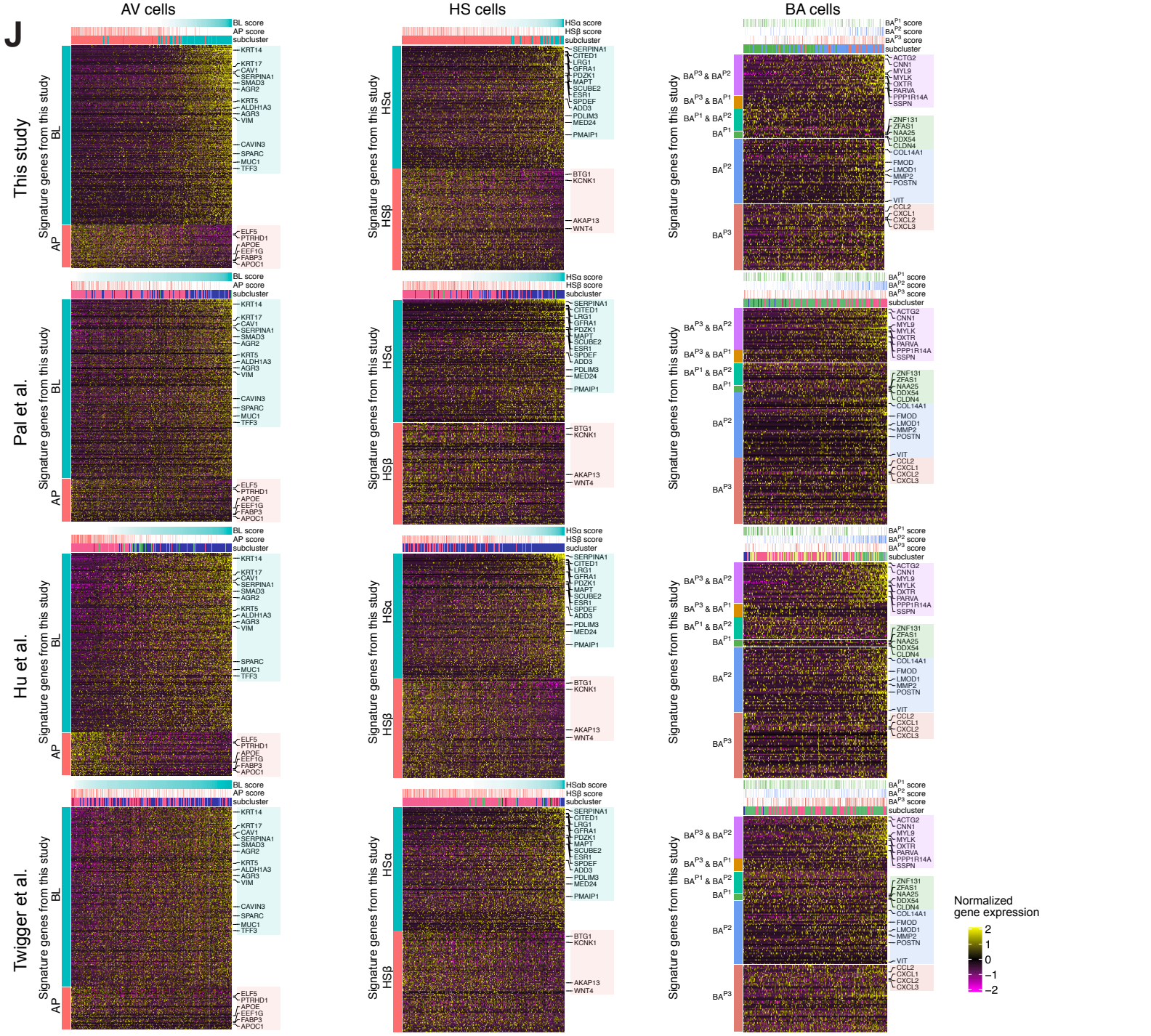


Figure S2. Diverse cell subtypes identified by scRNA-seq (related to Figure 2)

(A) UMAP visualization of the stromal subtypes. Fibroblast subtypes include F1, F2, and F3. Vascular/lymphatic cells include vascular endothelial cells (VE), lymphatic endothelial cells (LE), and pericytes (PE). Immune cells include myeloid cells (M), natural killer cells (NK), T cells (T), B cells (B), and plasma cells (P). Cells in gray indicate minor subclusters discarded as potential doublets due to unusually high gene counts and aberrant marker expression. To ensure that the subpopulations are generalizable across samples, minor subclusters representing only a limited subset of samples were merged with adjacent subclusters, as described in STAR Methods.

(B) Proportion of the fibroblast, vascular/lymphatic, and immune subtypes across 16 samples. Color code is identical to (A).

(C) Violin plots showing expression of markers associated with stromal subtypes.

(D) Heatmaps showing stromal subpopulation-specific gene expression signatures. (See Table S2.)

(E) UMAP visualization of subsets within T cells, myeloid cells, and plasma cells.

(F) GSEA plots of the degree of enrichment of AV, HS, and BA signatures in the indicated BL differential expression comparisons. Relative to AP cells, BL cells are enriched for HS and BA markers (top row; BL vs AP signature). The absolute value of the HS marker normalized enrichment score (NES) in the BL vs HS signature (2.80; middle row, middle panel) is higher than that of the BL vs AP signature (1.84; top row, middle panel), suggesting that the expression levels of HS markers in HS cells themselves are higher than in BL cells. The same is true for BA markers (2.79 vs 1.78) in the BL vs BA signature (bottom row, right panel) and BL vs AP signature (top row, right panel). Finally, the high NES values for the AV markers in the BL vs HS signature (2.51) and BL vs BA signature (2.64) suggest that BL cells are strongly associated with AV markers. Overall, these results show that BL cells are an AV subset that partially expresses HS and BA markers, albeit at a lower magnitude of expression. (Related to Figure 2F.)

(G) UMAP visualization of AV, HS, and BA cell subtypes before and after IEG regression and re-clustering (top). Correlation of pre- and post-regression log fold changes in differential gene expression analysis of the indicated cell subtypes (bottom; Pearson correlation analysis). BA cell subtypes identified post-regression are annotated by superscript P. In contrast to BA cells, AV and HS cell transcriptional signatures were largely unaltered by IEG regression. Of note, despite the differing BA clustering after regression, BAα cells were almost wholly assigned to the BA^{P3} cluster.

(H) Heatmaps showing gene expression signatures of BA subtypes after IEG regression and re-clustering. The main non-IEG genes distinguishing the pre-regression clusters (namely, contractility genes) also drove the clustering of the post-regression clusters, thereby highlighting both the centrality of contractile potential to BA diversity and the robustness of the clustering approach overall. (See Table S2.)

(I) UMAP of AV, HS, and BA cells identified in this study and by re-analysis of three publicly available scRNA-seq datasets (Hu et al., 2021; Pal et al., 2021; Twigger et al., 2022), colored by the indicated cell subtype signature score identified in this study (see Table S2) or by cell subclusters identified in each dataset using the analysis pipeline established in this study (see STAR Methods). Subclusters identified in the public datasets are assigned a unique color palette to emphasize their derivation independent of the present dataset's subtypes. The BL, AP, HSα, HSβ, BA^{P1}, BA^{P2}, and BA^{P3} cell subtype signatures defined in this study identified corresponding signature-positive cells as closely associated subpopulations in UMAP visualizations of all three published datasets.

(J) Heatmaps of cells in the three scRNA-seq datasets described in (I), showing gene expression patterns of cell subtype signatures identified in this study. Color codes for signature scores and subclusters are identical to (I). The genes within each of the subtype signatures

defined in this study were also coordinately expressed in the corresponding cells from each of the public datasets, providing additional support for the robustness of these subtype designations.

Supplemental Figure S3

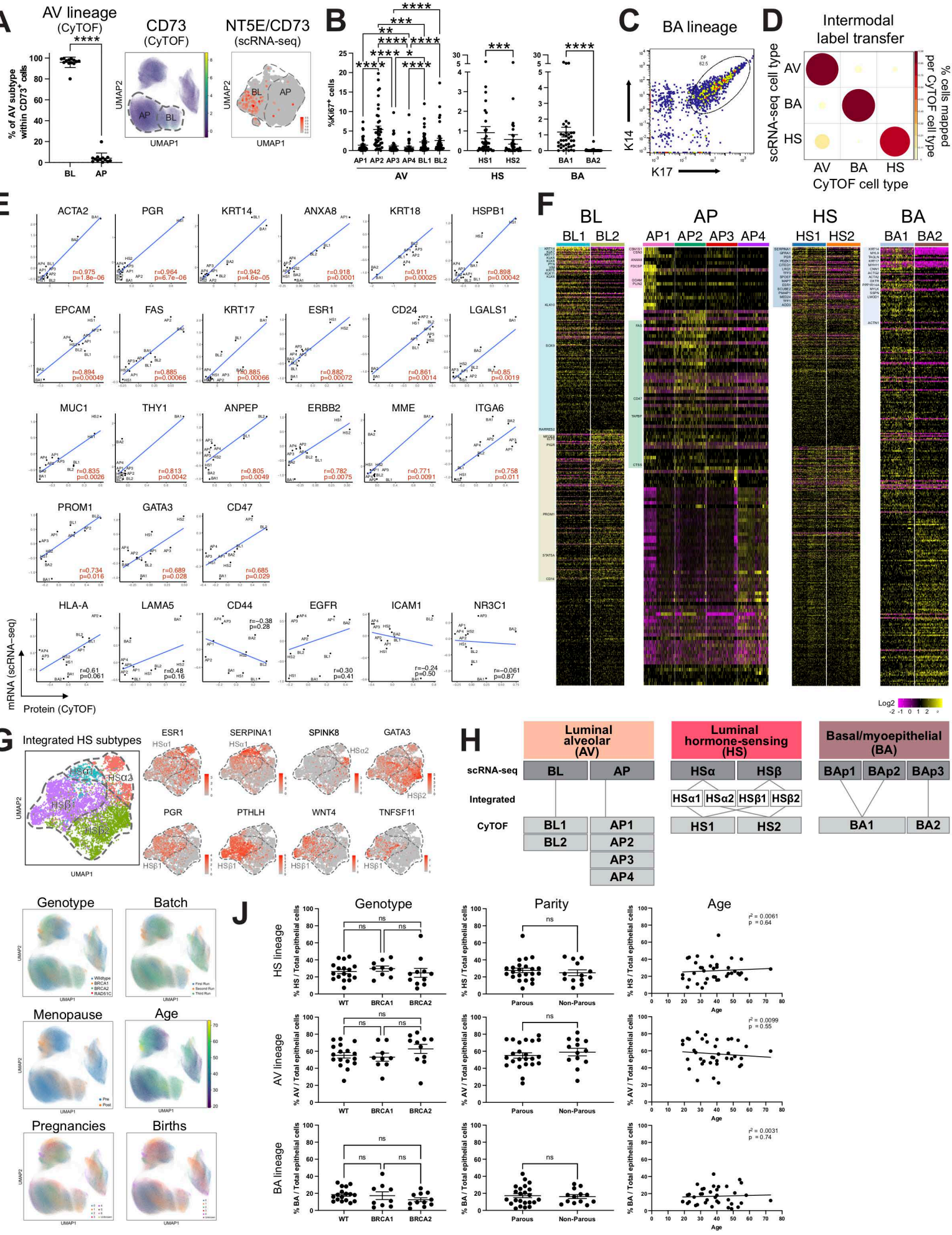


Figure S3. Diverse cell subtypes derived from CyTOF-based clustering analysis integrated with scRNA-seq data
(related to Figure 3)

(A) Percentage of $CD73^+$ cells in the BL or AP subtype in each sample by CyTOF, showing that $CD73^+$ cells are predominantly BL cells rather than AP cells (left; paired *t*-test). UMAP of CD73 protein level in CyTOF (middle) and mRNA level in scRNA-seq data (right). Error bars represent mean ± SEM.

(B) Percentage of $Ki67^+$ cells in each subtype. Non-parametric paired comparisons were applied between subtypes within each lineage, specifically Dunn's multiple comparisons test (AP1-4 and BL1-2) and Wilcoxon matched-pairs signed rank test (HS1 and HS2; BA1 and BA2). Error bars represent mean ± SEM.

(C) Biaxial plot of K14 and K17 levels in BA cells in a representative sample in CyTOF data.

(D) Dot plot of major epithelial cell type association across modalities after intermodal label transfer from CyTOF to scRNA-seq dataset. Size and color of dots represent the percentage of cells assigned to each cluster.

(E) Correlation plots of CyTOF markers comparing protein and mRNA levels in the integrated clusters. Plotted values are the mean-normalized and scaled pseudo-bulk expression values from CyTOF (x-axis) and scRNA-seq (y-axis).

(F) Heatmaps showing cluster-specific gene expression signatures in integrated data. (See Table S3.)

(G) UMAP of the four HS subtypes as defined by integrated scRNA-seq and CyTOF data (left) and mRNA expression patterns of example markers for each of the four HS subtypes (right). Notably, there were clear expression gradients of ER-regulated genes.

(H) Summary of cell subtype integration across scRNA-seq and CyTOF data.

(I) UMAPs of epithelial CyTOF data colored by clinical variables.

(J) Relative abundance of HS (top), AV (middle), and BA (bottom) cells within the epithelial compartment across genotypes (one-way ANOVA), parity (*t*-test), and age (simple regression) in CyTOF data (n=38 samples). Error bars represent mean ± SEM.

Supplemental Figure S4

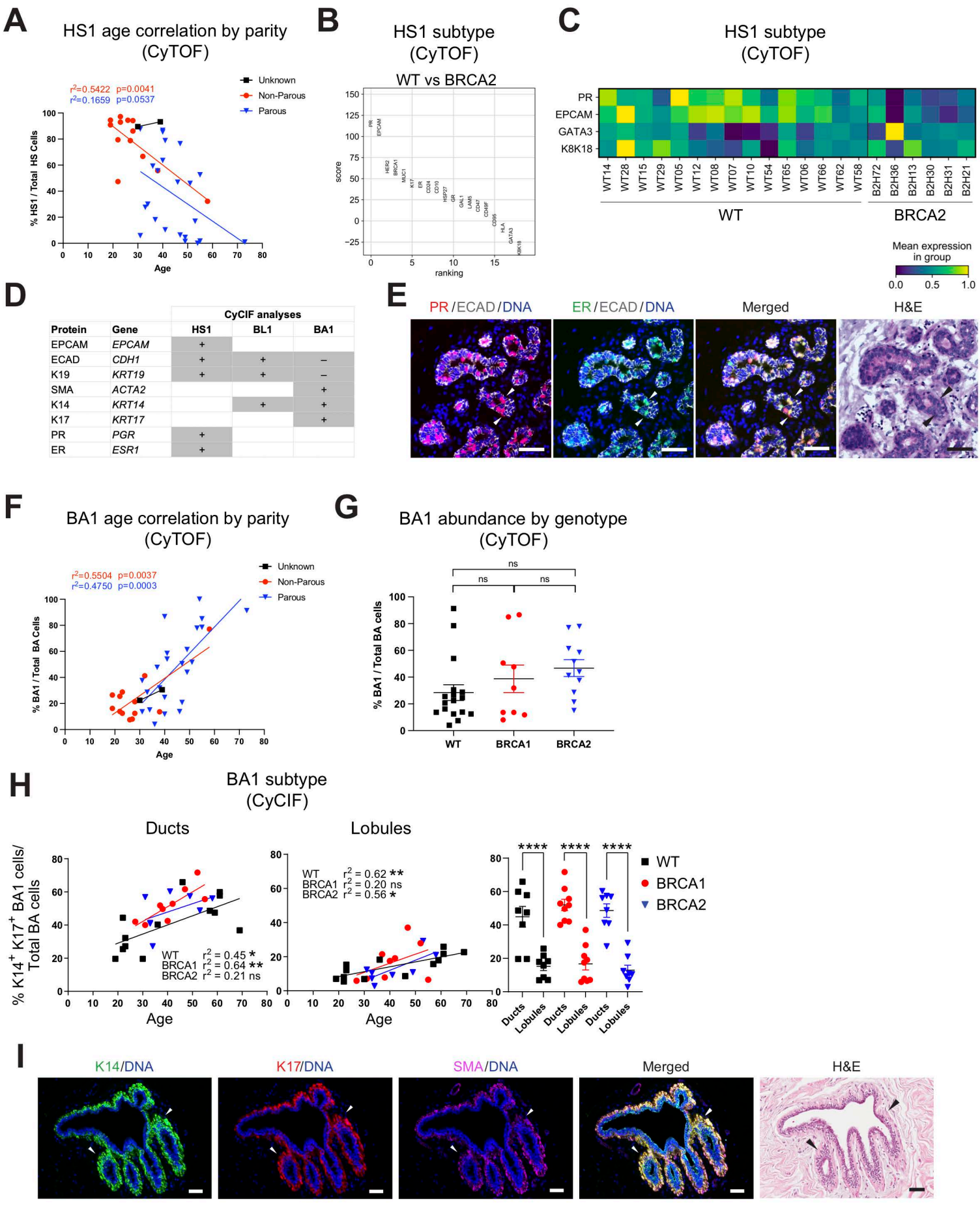


Figure S4. Correlation of HS and BA subtypes with clinical variables in CyTOF data (related to Figures 4 and 5)

(A) Relative abundance of HS1 cells within each parity group across age (simple regression analysis). Note that analysis of the effects of parity was confounded by a strong correlation between parity and age.

(B) Ranking of CyTOF markers in HS1 cells by fold difference between *BRCA2* noncarrier and carrier samples.

(C) CyTOF heatmap showing protein levels of PR in *BRCA2* noncarriers or carriers. All donors are premenopausal and have not undergone salpingo-oophorectomy or chemotherapy.

(D) CyCIF antibody panel.

(E) CyCIF staining of HS1 cells (ER^+ PR^+; arrowheads) with matching H&E staining. Scale bar=50 μm. (Related to Figure 4E.)

(F) Relative abundance of BA1 cells within each parity group across age (simple regression analysis). Note that the analysis of the effects of parity was confounded by a strong correlation between parity and age.

(G) Proportion of BA1 cells within the BA lineage across genotypes in the CyTOF data (noncarriers=17, *BRCA1* carriers=9, *BRCA2* carriers=11; analyzed by one-way ANOVA). Error bars represent mean ± SEM. (Related to Figure 5B.)

(H) Quantification of BA1 proportion in CyCIF stained breast sections (30 samples total, including noncarriers=13, *BRCA1* carriers=9, and *BRCA2* carriers=8). The percent of BA1 subtype ($K14^+$ $K17^+$ SMA^+ and [$K19^-$ and/or $ECAD^-$]) area out of total BA cell type (SMA^+ and [$K19^-$ and/or $ECAD^-$]) area within ducts or lobules was plotted against donor age within each genotype (left and middle panels) and plotted in aggregate by genotype for sections with sufficient representation of both ducts and lobules (right). Age correlations were analyzed by simple regression. Duct vs. lobule comparisons by genotype were analyzed by two-way ANOVA with Bonferroni correction.

(I) CyCIF staining of BA1 cells ($K14^+$ $K17^+$; arrowheads) with matching H&E staining. Scale bar=50 μm. (Related to Figure 5E.)

Supplemental Figure S5

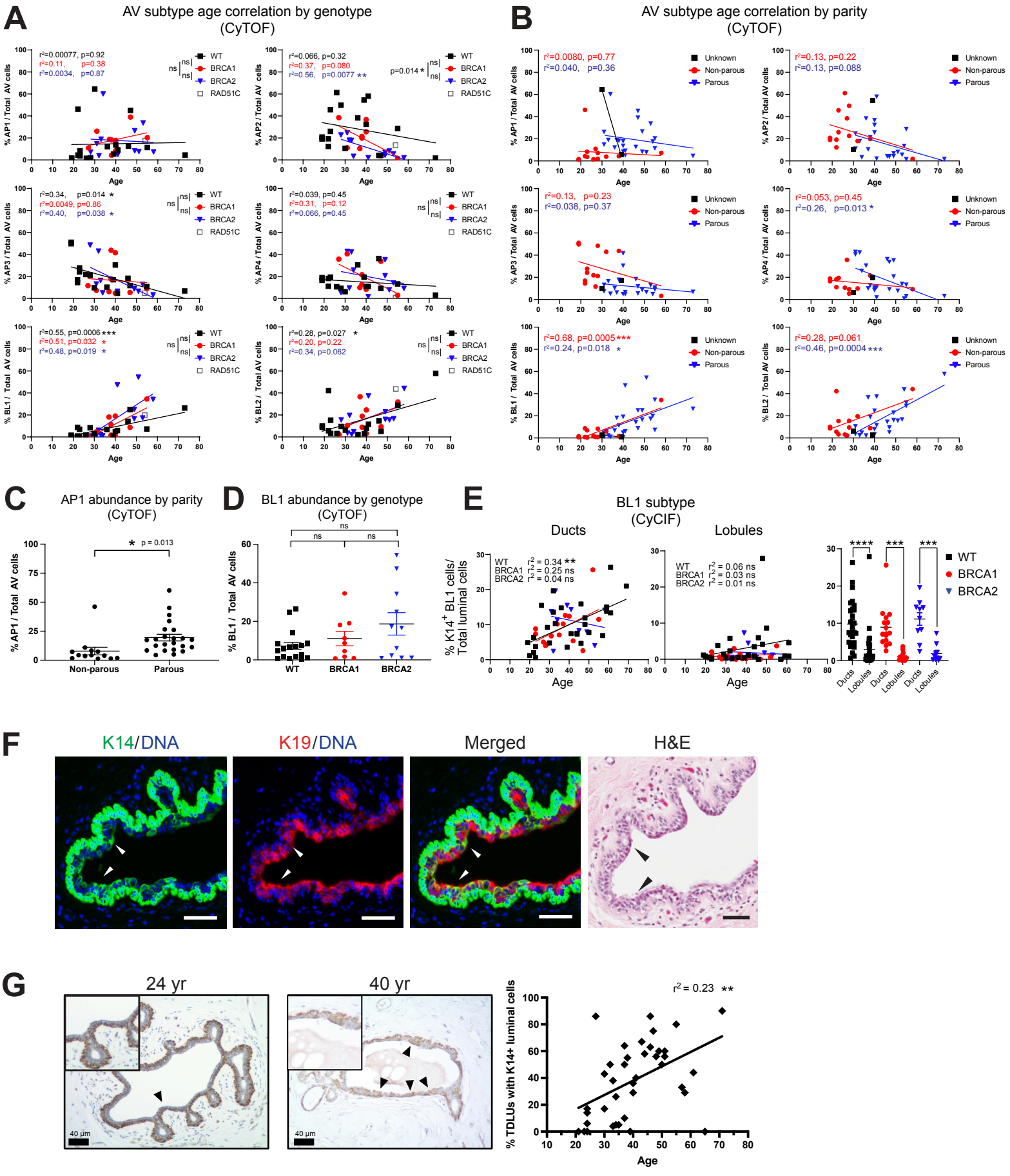


Figure S5. Correlation of AV subtypes with clinical variables in CyTOF data (related to Figure 6)

(A) Relative abundance of the six AV cell subtypes across age by genotype. Age correlations were analyzed by simple regression analysis, and genotype comparisons were analyzed by one-way ANOVA.

(B) Relative abundance of the six AV cell subtypes within each parity group across age. Simple regression analysis.

(C) Relative proportion of AP1 cells within the AV lineage in non-parous and parous women. Welch's *t*-test. Error bars represent mean ± SEM.

(D) Proportion of BL1 cells within the AV lineage across genotypes in the CyTOF data (noncarriers=17, *BRCA1* carriers=9, and *BRCA2* carriers=11; analyzed by one-way ANOVA).

(E) Quantification of BL1 cells in CyCIF stained breast sections (52 samples total, including noncarriers=27, *BRCA1*=15, *BRCA2*=10). The percent of BL1 cells ($K14^+$ and [$K19^+$ or $ECAD^+$]) out of total luminal cells ($K19^+$ or $ECAD^+$) within ducts or lobules was plotted against donor age within each genotype (left and middle panels) and plotted in aggregate by genotype for sections with sufficient representation of both ducts and lobules (right). Age correlations were analyzed by simple regression. Duct vs. lobule comparisons by genotype were analyzed by two-way ANOVA with Bonferroni correction. Error bars represent mean ± SEM. (Related to Figure 6D.)

(F) CyCIF staining of BL cells ($K14^+$ $K19^+$; arrowheads) with matching H&E staining. Scale bar=50 μm. (Related to Figure 6E.)

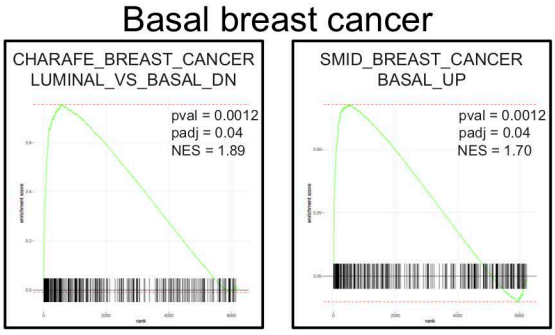
(G) K14 IHC staining of ductal structures from two representative samples (left). Arrow heads indicate $K14^+$ BL1 cells in the luminal layer. Quantification of percent terminal ducts in TDLUs with $K14^+$ BL1 cells across 39 samples (right). An average of 13 TDLUs (range 5-44) per sample were scored by a breast pathologist in a blinded fashion.

Supplemental Figure S6

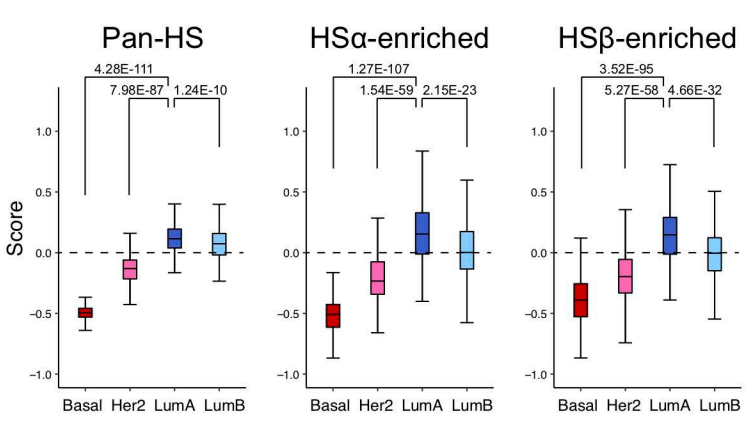
A

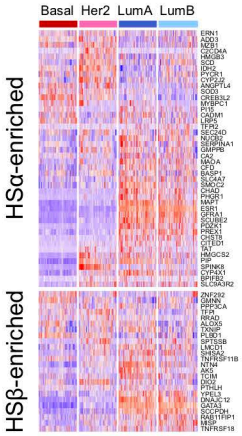
Cell type/subtype signature	Pairwise comparison						Aggregate comparison
	Basal vs Her2	Basal vs LumA	Basal vs LumB	Her2 vs LumA	Her2 vs LumB	LumA vs LumB	
Pan-AV	4.42E-49	4.76E-105	1.78E-101	1.19E-78	1.56E-81	1.70E-14	1.42E-191
BL-enriched	2.87E-08	8.66E-47	6.87E-56	1.81E-13	9.70E-25	3.07E-10	2.52E-73
AP-enriched	1.11E-10	3.98E-48	4.13E-19	9.49E-78	9.46E-46	1.21E-12	3.98E-103
Pan-HS	4.38E-73	4.28E-111	5.04E-106	7.98E-87	9.34E-65	1.24E-10	7.01E-195
HSα-enriched	2.66E-45	1.27E-107	4.76E-97	1.54E-59	2.14E-30	2.15E-23	1.98E-166
HSβ-enriched	3.99E-20	3.52E-95	2.33E-72	5.27E-58	1.56E-25	4.66E-32	3.57E-147
Pan-BA	6.68E-16	3.49E-35	4.03E-62	3.23E-04	2.75E-26	2.81E-23	2.17E-77
BA^{P2}-enriched	5.90E-04	2.34E-01	2.56E-18	1.69E-10	5.52E-08	3.72E-55	3.21E-55
BA^{P3}-enriched	3.47E-01	3.48E-01	8.18E-15	9.29E-01	4.19E-25	2.89E-41	7.42E-46

B

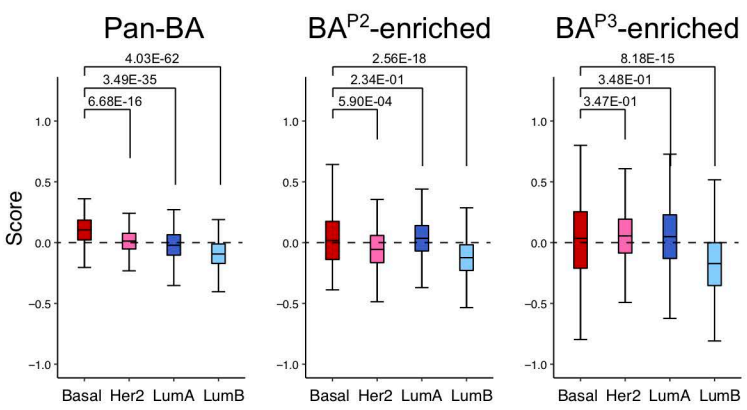


C





D



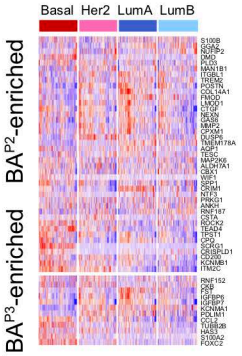


Figure S6. Association of BL transcriptomic signature with basal-like breast cancer (related to Figure 6)

(A) P-values of aggregate comparisons (Kruskal-Wallis test) and pairwise comparisons (Wilcoxon rank sum test) between each cell type/subtype signature in scRNA-seq data and transcriptomic profiles of breast cancer subtypes in METABRIC (basal-like, HER2+, Luminal A, and Luminal B). (Related to Figure 6F, S6C, and S6D.)

(B) Basal-like breast cancer pathway-related gene sets enriched in the BL signature in scRNA-seq data as identified by GSEA. (See Table S2.)

(C-D) Box plots (left) showing associations between scRNA-seq signatures of HS (C) or BA (D) cell types/subtypes and transcriptomic profiles of breast cancer subtypes in METABRIC. Aggregate and pairwise comparisons among tumor subtypes were analyzed by Kruskal-Wallis and Wilcoxon rank sum tests, respectively. Boxes represent the first, second (median), and third quartiles; whiskers span 1.5x interquartile range from the first/third quartile. Heatmaps (right) showing gene signature expression across cancer subtypes. Tumors were randomly downsampled to match the subtype with the smallest sample size. The filtered BA^{P1}-enriched signature contained too few genes and was not used in this analysis. (See Table S4.)

Supplemental Figure S7

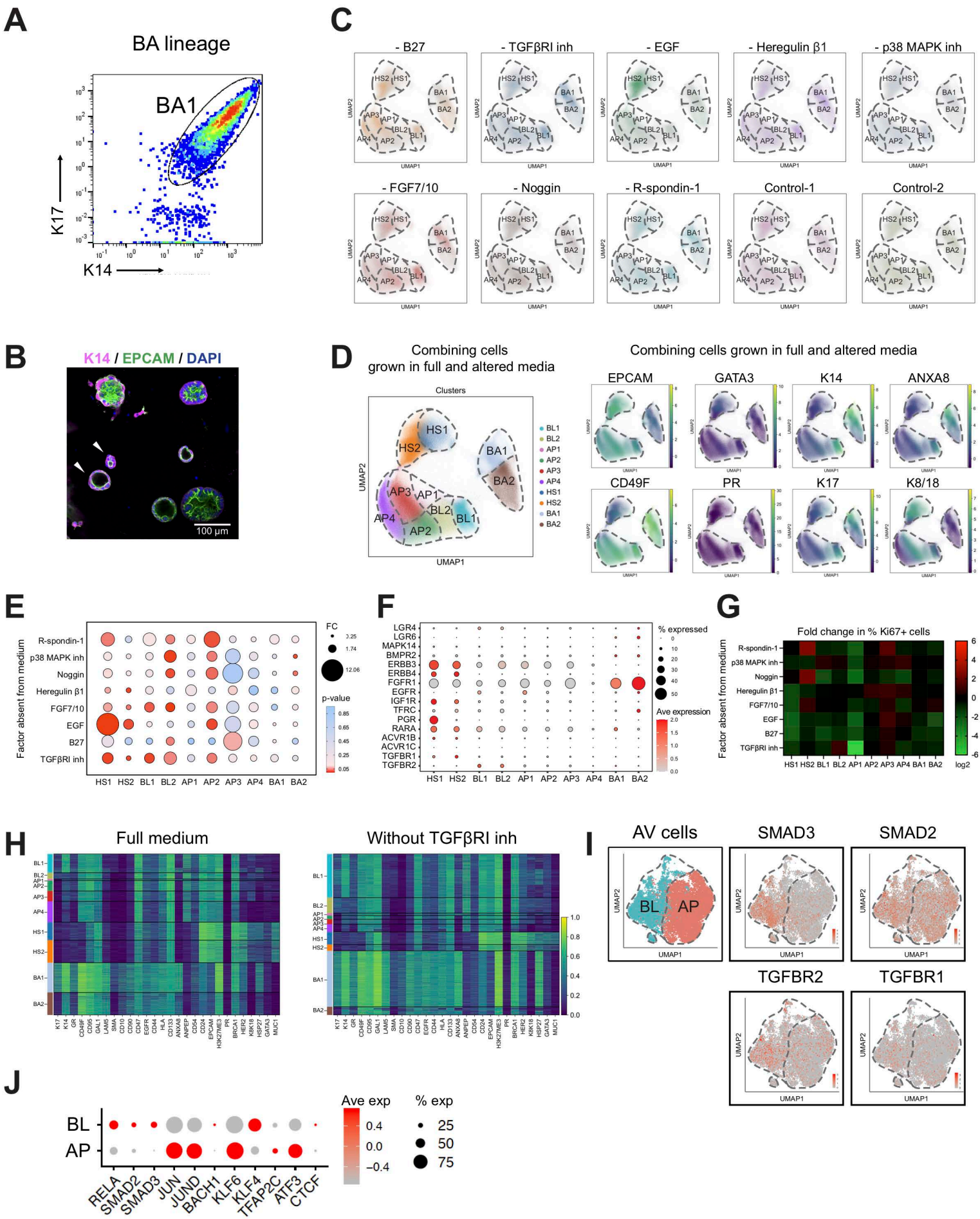


Figure S7. Organoid cultures as a model system to identify signaling pathway dependencies in cell subtypes
(related to Figure 7)

(A) CyTOF biaxial plot showing K14 and K17 levels in BA cells ($EPCAM^{lo}$ $CD49F^{hi}$) in a representative organoid culture. BA1 cells ($K14^{+}$ $K17^{+}$) are circled.

(B) Representative confocal immunofluorescence microscopy image of an organoid culture stained for K14 and EPCAM and counterstained with DAPI. Arrowheads indicate luminal-type organoids (Rosenbluth et al., 2020) that express K14, showing preservation of BL1 cells *in vitro*.

(C) CyTOF UMAPs showing all cells in organoid cultures (n=9) grown in full medium or the indicated altered media. In each panel, cells grown under the indicated condition are colored, and other cells are gray.

(D) UMAP displaying CyTOF analyses of 9 organoid cultures. Both cells grown under full media conditions and cells grown in altered media are graphed. Cells from organoid cultures were projected onto the clusters defined in the UMAP of fresh tissues (Figure 3A) based on correlation analyses of protein expression patterns (left panel). UMAPs of markers associated with epithelial cell subtypes (right panels).

(E) Average fold changes in the relative proportions of cell subtypes in organoid cultures in the absence of the indicated media components, colored by p-value (*t*-test).

(F) Expression levels of the indicated receptors for each of the factors modulated in the organoid medium are shown from the integrated dataset for each subtype.

(G) The fold change in percentage of $Ki67^{+}$ cells (defined as >4 on arcsinh scale in CyTOF data) in organoids grown in the indicated conditions relative to organoids grown in full medium.

(H) Single-cell heatmaps of CyTOF marker expression in epithelial subtypes in a representative organoid culture with or without removal of TGFβRI inhibitor.

(I) UMAP visualization of *SMAD3*, *SMAD2*, *TGFBR2*, and *TGFBR1* mRNA levels in BL and AP cells within the AV lineage.

(J) Expression levels of transcription factors listed in Figure 7H in the scRNA-seq dataset.